Reading and lexicalization in opaque and transparent orthographies: Word naming and word learning in English and Spanish

Kwok, RKW, Cuetos, F, Avdyli, R & Ellis, AW

Original citation & hyperlink:
Kwok, RKW, Cuetos, F, Avdyli, R & Ellis, AW 2016, 'Reading and lexicalization in opaque and transparent orthographies: Word naming and word learning in English and Spanish' *Quarterly Journal of Experimental Psychology*, vol 70, no. 10, pp. 2105-2129
https://dx.doi.org/10.1080/17470218.2016.1223705

DOI 10.1080/17470218.2016.1223705
ISSN 1747-0218
ESSN 1747-0226

Publisher: Taylor and Francis

Reading and lexicalisation in opaque and transparent orthographies: word naming and word learning in English and Spanish

Journal:	*Quarterly Journal of Experimental Psychology*
Manuscript ID	QJE-STD 15-335.R2
Manuscript Type:	Standard Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Kwok, Rosa; Coventry University, Psychology Cuetos, Fernando; University of Oviedo, Psychology Avdyli, Rrezarta; University of Oviedo, Psychology Ellis, Andy; University of York, Psychology
Keywords:	word naming, word learning, word length, orthography, transparency, grain size theory

Reading and lexicalisation in opaque and transparent orthographies: word naming and word learning in English and Spanish

Rosa Kit Wan Kwok[1], Fernando Cuetos[2],

Rrezarta Avdyli[2], Andrew W. Ellis[3]

[1]*Department of Psychology, Coventry University, Coventry, UK*

[2]*Department of Psychology, University of Oviedo, Oviedo, Spain*

[3]*Department of Psychology, University of York, York, UK*

Running head: Word learning in Spanish and English

Correspondence should be addressed to:

Fernando Cuetos, Facultad de Psicología, Plaza Feijoo, s/n, 33003 Oviedo, Spain. Email: fcuetos@uniovi.es. We thank Robert Davies for his comments and his suggestions regarding the presentation of the code.

ABSTRACT

Do skilled readers of opaque and transparent orthographies make differential use of lexical and sublexical processes when converting words from print to sound? Two experiments are reported which address that question, using effects of letter length on naming latencies as an index of the involvement of sublexical letter-sound conversion. Adult native speakers of English (Experiment 1) and Spanish (Experiment 2) read aloud four- and seven-letter high frequency words, low frequency words and nonwords in their native language. The stimuli were interleaved and presented ten times in a first testing session and ten more times in a second session 28 days later. Effects of lexicality were observed in both languages, indicating the deployment of lexical representations in word naming. Naming latencies to both words and nonwords reduced across repetitions on day 1, with those savings being retained to day 28. Length effects were, however, greater for Spanish than English word naming. Reaction times to long and short nonwords converged with repeated presentations in both languages, but less in Spanish than in English. The results support the hypothesis that reading in opaque orthographies favours the rapid creation and use of lexical representations while reading in transparent orthographies makes more use of a combination of lexical and sublexical processing.

Keywords: reading, word naming, word learning, word length, orthography, transparency, grain size theory, English, Spanish.

Consider the English words *boot*, *coot*, *loot*, *root* and *shoot*. They share the same -*oot* ending and are pronounced the same way (i.e., they rhyme). They contrast, however, with *foot* and *soot* which have the same ending but a different pronunciation. There are more words in the *boot* / *coot* / *loot* family than in the *foot* / *soot* family, so the pronunciation in *boot* and *coot* is said to be regular while the pronunciation in *foot* and *soot* is deemed irregular or exceptional (Coltheart, Davelaar, Jonasson, & Besner, 1977; Plaut, McClelland, Seidenberg, & Patterson, 1996). English contains a large number of irregular or exception words, including many of the most common words in the language (e.g., *have, once*, *said*, *some*, *two*, *was* and *women*). Given that inconsistency, the only reliable way to read words in English is to recognise them as familiar sequences of letters and retrieve their pronunciations from lexical memory as whole units.

In Spanish, the situation is different. Anyone familiar with the way that Spanish letters and letter combinations are pronounced can read the Spanish equivalents of *boot*, *foot*, *some* and *women* (*bota*, *pie*, *algunos*, *mujeres*) with confidence, knowing that the pronunciations assigned to those words will be correct. Unfamiliar Spanish words can also be read with confidence using the letter-sound mappings employed in familiar words. That is why Spanish is said to have a *transparent orthography* while English has an *opaque orthography*.

The questions addressed here are whether the transparency of Spanish and the opacity of English affect the way skilled readers of the two languages convert words from print to sound and the way that new words are learned in the two languages. In principle, the pronunciations of all Spanish words could be assembled *de novo* using sublexical letter-sound correspondences each time they are read. There are good reasons, however, for believing that this is not what happens. If the pronunciation of a common (high frequency) Spanish word like *malo* (bad) was assembled from sublexical units each time it was pronounced, there would be no reason to expect that the time from seeing *malo* on a page or computer screen to reading it aloud (i.e., its naming latency) would be any different from the time required to read aloud a low frequency word like *mago* (wizard) or even a nonword like *maco*. Faster naming of high than low frequency Spanish words (a *frequency effect*) has, however, been observed many times (Alvarez, Carreiras, & Taft, 2001; Carreiras, Alvarez, & De Vega, 1993; Carreiras, Vergara, & Barber, 2005; Davies, Barbón, & Cuetos, 2013; González-Nosti, Barbón, Rodríguez-Ferreiro, & Cuetos, 2013; Perea & Carreiras, 1998; Perea, Carreiras, & Grainger, 2004). Faster naming of familiar words than nonwords (a *lexicality effect*) has been observed in Spanish children, with the effect becoming stronger as the children grow older and more words become familiar (Castejón, Rodríguez-Ferreiro, & Cuetos, 2013; Cuetos & Suárez-Coalla, 2009; Davies, Rodríguez-Ferreiro, Suárez, & Cuetos, 2013). Lexicality and frequency effects have also been reported in Italian, another language with a transparent orthography (Colombo, Pasini, & Balota, 2006; Pagliuca, Arduino, Barca, & Burani, 2008; Paizi, Burani, & Zoccolotti, 2010; Zoccolotti, De Luca, Di Filippo, Judica, & Martelli, 2009).

Readers of Spanish and Italian develop lexical representations that allow faster naming of familiar words (especially high frequency words) than nonwords. We can therefore exclude the possibility that skilled reading in transparent orthographies is mediated entirely by sublexical letter-sound correspondences. It remains possible, however, that because lexical and sublexical reading processes generate the same pronunciations in a transparent orthography, skilled readers of Spanish routinely employ a combination of lexical and sublexical processes when reading familiar words, In English, conflict and interference between the results of lexical and sublexical processing, particularly for high frequency words, may result in the development of a reading system in which skilled readers make more exclusive use of lexical processing. Similarly, when new words are learned in the two languages, we might expect English readers to switch rapidly from sublexical to lexical processing while Spanish readers continue to use both modes of conversion from orthography to phonology.

Orthographic transparency and the 'grain size' theory

The idea that reading an opaque orthography results in more emphasis being placed on lexical mappings between orthography and phonology while reading a transparent orthography utilises lexical and sublexical mappings together was developed in the 'grain size' theory of Ziegler, Perry, Jacobs, and Braun (2001) and Ziegler and Goswami (2005) who argued that an influence of sublexical letter-sound conversion on word naming (reading aloud) should be detectable through an impact of letter length on naming latencies. The idea is that sublexical letter-sound conversion of the sort used, for example, to pronounce a new word for the first time, is a serial process that operates from left to right so requires more time to convert a long sequence of letters from print to sound than a short one (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Rastle & Coltheart, 1994; Rastle, Kinoshita, Lupker, & Coltheart, 2003). In contrast, lexical pronunciation involves a more parallel conversion of letter sequences to sound and therefore shows smaller length effects. Evidence in support of this proposal comes from studies that have reported stronger effects of length on naming latencies for unfamiliar words (nonwords) than familiar words in English (e.g., Ellis et al., 2009; Hogaboam & Perfetti, 1978, Mason, 1978; Weekes, 1997). In Weekes' (1997) study, for example, English nonwords composed of three, four, five or six letters (e.g., *bot*, *leck*, *grite*, *stetch*) were interleaved with high frequency words (e.g., *bed*, *land*, *green*, *spring*) and low frequency words (e.g., *bog*, *loin*, *grunt*, *swerve*) matched on initial letters and phonemes. The influence of length on naming latencies was strong and highly significant for nonwords, reduced for low frequency words and non-significant for high frequency words.

Coltheart et al. (2001, p. 239) demonstrated that their computational model of visual word recognition (the DRC model) was able to simulate the reported differences in length-sensitivity between words and nonwords. In the DRC model, nonwords are converted from print to speech using grapheme-phoneme (letter-sound) conversion rules that operate from left to right and therefore take more time (cycles) to 'pronounce' a long nonword than a short one. Words, in contrast, have visual representations in an orthographic input lexicon and speech-based representations in a phonological output lexicon. Those representations allow words to be converted from print to sound more rapidly and more wholistically than nonwords.

Ziegler et al. (2001) sought to test the notion embodied in their grain size theory that length effects – the hallmark of sublexical processing – would be more evident for words in a transparent orthography (German) than an opaque orthography (English). The stimuli in their study were words and nonwords that are similar or identical in the two languages (e.g., words: *sand* / *Sand*, *storm* / *Sturn*; nonwords: *fot* / *Fot*, *ploar* / *Plohr*). In line with predictions, analysis of naming latencies showed an interaction between length and lexicality (stronger length effects for nonwords than words) and an interaction between length and language (larger overall length effects for German than English). There was no firm evidence, however, that the differences in length effects between German and English were specific to words (where length effects were expected to be greater in German than English) rather than nonwords (where length effects were expected in both languages).

Perry and Ziegler (2002) explored the grain size theory in computational terms by comparing length effects for words and nonwords in the original DRC model of English word recognition (Coltheart et al., 2001) and in versions of that model that were adapted for the German language (based on Ziegler, Perry, & Coltheart, 2000). The English and German models were tested on words and nonwords ranging in length from three to six letters. In the critical simulations, the German DRC model was adjusted to make the relative speeds of the lexical and sublexical routes more equal than was the case for the English model. That is, the simulations were based on a version of grain size theory in which reading in both opaque and transparent orthographies employs large grain (lexical) and small grain (sublexical) mappings between orthography and phonology, but the pronunciation of familiar words in opaque orthographies is based predominantly on lexical representations while the pronunciation of familiar words in transparent orthographies is based on more equal contribution from lexical and sublexical mappings. As would be expected, Perry and Ziegler (2002) found stronger length effects for words in the German DRC model than in the English model.

Word learning, lexicalisation and the grain size theory

The present experiments investigated the effects of letter length on naming latencies for high and low frequency words in English (Experiment 1) and Spanish (Experiment 2). The words contained either four or seven letters. They were presented ten times in a first testing session and ten more times in a second session 28 days later. On the basis of grain size theory, we predicted stronger length effects in Spanish than English word naming, particularly for less familiar (low frequency) words. We expected that those effects would diminish as a result of repetition as the words benefited from the effects of repetition.

The familiar words were interleaved with four- and seven-letter nonwords in each block of trials. The rationale for including nonwords in the two experiments was to enable us to compare the process of word learning in the two languages and, in particular, to investigate the speed with which the establishment of lexical representations caused reading to switch from sublexical, length-sensitive processing to a lexical mode of processing in which length effects are reduced. Maloney, Risko, O'Malley, and Besner (2009) presented Weekes' (1997) nonword stimuli to adult participants four times in different random orders with instructions to read the nonwords aloud as quickly as possible as they appeared on the computer screen. Naming latencies became faster across blocks while the effect of length diminished. Maloney et al. (2009) argued that the reduction in RTs to nonwords and the convergence of RTs to shorter and longer items reflected the creation of lexical representations and a consequent switch from sublexical to lexical processing. They proposed that three or four presentations was enough for skilled readers of English to create functioning lexical representations for novel words that could sustain lexical reading.

Kwok and Ellis (2015) replicated and extended the Maloney et al. (2009) study. Kwok and Ellis (2015) created nonwords that contained either four letters or seven letters. In their Experiment 1, the nonwords were presented ten times to skilled adult readers across ten blocks of trials. Naming latencies declined markedly over the first four or five presentations, reaching an asymptote around block six or seven. The difference in reaction times (RTs) to short and long nonwords also reduced with repeated presentations, becoming non-significant around the same time as the RTs asymptoted (see also Kwok & Ellis, 2014). In Experiment 3 of Kwok and Ellis (2015), the participants returned after seven days and repeated the process of reading the nonwords ten times across ten blocks of trials. A small length effect was observed in block one of day 7, but after that RTs were fast and there was no difference in naming speeds between short and long items; that is, participants showed good retention of word learning across seven days. Like Maloney et al. (2009), Kwok and Ellis (2015) argued that a shift from slow, serial, sublexical reading of unfamiliar nonwords to faster, more parallel reading occurred as the result of lexical representations being established over the first four to six presentations (cf. Salassoo, Shiffrin, & Feustel, 1985). The lexical representations then survived largely intact over a period of seven days (see also Kwok & Ellis, 2014).

Suárez-Coalla and Cuetos (2016) reported a similar study of novel word learning in Spanish. They presented real but unfamiliar Spanish words six times in a first testing session and six more times in a second session a month later. The stimuli were either short (four to five letters) or long (seven to eight letters). Two groups of participants received prior training on the novel words, but a third group received no pre-training. Naming latencies for the group that received no pre-training reduced across blocks on day 1, as did the difference in RTs to short and long items.

To summarise, the experiments that follow evaluated the grain size theory of reading in terms of the following predictions:

1. Greater involvement of sublexical processing in reading familiar words in transparent compared with opaque orthographies will be reflected in greater length effects in Spanish than English word reading.

2. Word learning will be characterised by a faster and more complete switch from sublexical to lexical processing in English than in Spanish. That will be reflected in faster convergence of RTs to short and long nonwords in English than Spanish.

In addition, we expected that naming latencies for both English and Spanish words would reduce across blocks as a result of repetition priming (Scarborough, Cortese, & Scarborough, 1977). Given the remarkably long-lasting effects of repetition priming for familiar stimuli (e.g., Cave, 1997; Wiggs, Weisberg, & Martin, 2006), we expected that the gains in naming latencies to words over the first testing session would be sustained over the 28-day retention interval. On the basis of the results reported by Kwok and Ellis (2014, 2015) and Suárez-Coalla and Cuetos (2016), we also expected that learning effects for nonwords seen on day 1 would be retained to day 28. These various predictions were tested in two separate experiments and in combined analyses of the results across experiments.

EXPERIMENT 1: REPEATED PRESENTATION OF HIGH FREQUENCY WORDS, LOW FREQUENCY WORDS AND NONWORDS ACROSS 28 DAYS IN ENGLISH

Experiment 1 involved the repeated presentation of interleaved high frequency words, low frequency words and nonwords to native speakers of English in two testing sessions 28 days apart. Naming latencies were measured as the interval between a stimulus appearing on a computer screen and the onset of the spoken response. Theoretical interest lay in the relative effects of length on naming latencies for the two classes of word and the nonwords, the extent to which RTs converged for short and long stimuli and the persistence of training / repetition effects over a 28-day retention interval. The results will also be used as a point of comparison for Experiment 2.

Method

Participants

The participants were 25 undergraduate students of the University of York, UK (13 female, 12 male) with a mean age of 20.08 years (S.D. = 2.68; range 18 - 31). All were native speakers of English with normal or corrected-to-normal vision and no history of reading or language problems. Participants received either course credit or a small payment in return for their participation. The experiment was approved by the Ethics Committee of the Department of Psychology, University of York.

Materials

The experimental stimuli were 24 high frequency words, 24 low frequency words and 24 nonwords. Within each set, 12 items contained 4 letters and one syllable while 12 contained 7 letters and two syllables. The short and long high frequency words, low frequency words and nonwords were matched on initial letters and phonemes. Twelve different onsets were used to make the items as distinct as possible. To optimise voice key activation, none of the stimuli began with a voiceless fricative (‘f’, ‘s’, ‘sh’ or ‘th’).

Frequency measures were taken from SUBTLEX (Brysbaert & New, 2009) which is based on the subtitles of English films and television programmes. They are expressed in Table 1 as Zipf values (log10 of the frequency of each word per billion words of text; van Heuven, Mandera, Keuleers, & Brysbaert, 2014).

The nonwords were pronounceable letter strings generated by the WordGen program (Duyck, Desmet, Verbeke, & Brysbaert, 2004) and based on the CELEX and Lexique databases (Baayen et al., 1993, 1995; New, Pallier, Brysbaert, & Ferrand, 2004). They were a different set from those used by Kwok and Ellis (2014, 2015). The nonwords were matched to the real words on letter length, syllable length, initial letters and phonemes, and mean log bigram frequency from WordGen. The experimental stimuli and their values on the matching variables are shown in Table 1. Eighteen additional high frequency words, low frequency words and nonwords (6 of each) were selected for use in practice trials.

(Table 1 about here)

Procedure

Participants were tested individually. After completing a consent form, session 1 began with participants being given practice on the task. This involved reading 18 items (six high-frequency words, six low-frequency words and six nonwords, with half the stimuli of each type containing four letters and half seven letters). The experiment then began with the 72 experimental stimuli being presented in a random order in block one (12 short high frequency words, 12 long high frequency words, 12 short low frequency words, 12 long low frequency words, 12 short nonwords and 12 long nonwords).

Participants were seated approximately 60 cm from a computer screen on which the words and nonwords were displayed in black, lower case letters on a white background. The stimuli were presented in 18-point Times New Roman font. Each trial consisted of a centrally-presented fixation cross displayed for 1,000 ms, followed by the word or nonword stimulus for 2,000 ms then a blank screen for 1,000 ms before the next trial began. Participants were instructed to read each item aloud as quickly and as accurately as possible. The 72 stimuli were presented once in a random order. Participants were informed when a block was complete and pressed the space bar on a computer keyboard to initiate the next block when they were ready to continue. This process was repeated across ten blocks with the stimuli being presented in a different random order in each block. Participants wore headphones with a high-sensitivity microphone connected to a voice key that was linked to the computer. Presentation of the stimuli and recording of naming latencies was controlled by E-prime experiment generator software (version 1.2; Schneider, Eschman, & Zuccolotto, 2002). The experimenter noted any trials in which the participant misread a nonword, hesitated or made a false start or other form of error. No feedback was given at any point.

Participants returned 28 days later for a second session which repeated the practice items and the 10 blocks of experimental stimuli using the same procedure as on day 1.

Results

Naming errors, hesitations and failures to activate the voice key were removed from the analysis along with RTs less than 100 ms or long than 2.5 SDs above the mean (defined separately for each participant in each block and for each length after removal of RTs less than 100 ms). Naming errors, hesitations and failures to activate the voice key occurred on 80 trials (0.2% of the total). An additional 40 RTs were removed at the stage of RT trimming (0.1%), leaving 35,880 RTs for analysis (99.6% of the total). Table 2 (Supplemental materials) shows the accuracy and RT results for correct, trimmed responses. Accuracy never fell below 98% correct for any stimulus type in any block of trials and was at ceiling on day 28. For that reason, we will confine our statistical analysis to the RT data.

Naming latencies (RTs).

Figure 1 shows the pattern of RTs for correct, trimmed responses to high frequency words, low frequency words and nonwords across days and blocks. The RT data were analysed using linear mixed effects modelling (LME). LME methods analyse all the available data and do not rely on averaging across participants or items. They allow differences in the baseline performance among participants and items (*random effects*) to be separated from the effects of the predictor variables (*fixed effects*) (Baayen, Davidson, & Bates, 2008). The analyses were conducted in R using the lme4 (version 1.1-11; Bates, Maechler, & Bolker, 2012) and lmerTest packages (version 2.0-30; Kuznetsova, Brockhoff, & Christensen, 2015). We used the lmerTest package (Kuznetsova et al., 2015) to calculate *p* values using Satterthwaite approximations to determine degrees of freedom.

(Figure 1 about here)

A series of analyses models were used to address different questions. The predictors (fixed effects) were Day (1 vs 28), Blocks (1 to 10), Length (4 vs 7 letters) and Stimulus type (high frequency words, low frequency words and nonwords). The factor of Stimulus type was used to extract two contrasts, Lexicality (high frequency words vs nonwords) and word Frequency (high vs low frequency words). We present the results in terms of the effects involving Day, Blocks, Lexicality, Frequency and Length.

RTs were log-transformed to reduce skew. We followed the recommendations of Barr, Levy, Scheepers, and Tily (2013) by estimating fixed effects in models that included random effects terms corresponding both to differences between participants or items in overall speed of responses elicited (random intercepts) and to differences between participants or items in the slopes of the effects of the predictor variables (random slopes). We used the likelihood ratio test (Barr et al., 2013; Pinheiro & Bates, 2000) to assess whether the inclusion of fixed or random effects was warranted by superior model fit to data. That is, we included as many slopes as were found to be warranted. Given that the full model with all the within-subject factors as random slopes did not converge, each of the final models incorporated random intercepts for both participants and items effects and by-participant random slopes for Blocks and Length. The data for all the analyses reported here are provided in Appendices 1 to 8 (Supplemental materials) while the R syntax and report of model selection for each analysis is provided in Appendix 9 (Supplemental materials). In the text we present summaries of the final models (i.e., the models that best fitted the data according to likelihood ratio test).

The data for the first three analyses (Models 1 to 3) are presented in Appendix 1 and the R syntax in Appendix 9 (Supplemental materials).

Model 1. Global analysis of RTs across days 1 and 28.

We began with a global analysis of RTs investigating the effects of Day (1 vs 28), Blocks (1 to 10), Lexicality (high frequency words vs nonwords), Frequency (high vs low frequency words) and Length (4 vs 7 letters). The results are shown in full in Table 3.

(Table 3 about here)

There were significant effects of Day (faster RTs on day 28 than day 1) and Blocks (RTs reducing across blocks), with a significant Day x Blocks interaction reflecting a larger change in RTs across blocks in day 1 compared with day 28 (see Figure 1. There was also a significant interaction of Day, Lexicality and Length, reflecting a difference between the length effects for nonwords and high frequency words that was greater on day 1 than day 28. A significant effect of Lexicality (faster overall RTs to high frequency words than nonwords) combined with Length in significant two-way (Lexicality x Length) and three-way (Blocks x Lexicality x Length) interactions. These interactions can be summarised by saying that across days 1 and 28, the difference in RTs between high frequency words and nonwords reduced across blocks and was greater for long than short items, with the impact of length and lexicality declining across blocks. There was no significant effect of word Frequency and no significant interactions involving Frequency. The results were investigated further through separate analyses of the data from day 1 (Model 2) and day 28 (Model 3).

Model 2. Effects of Lexicality, Frequency, Blocks and Length on day 1

Model 2 analysed the data for day 1 only, looking at the effects of Blocks, Lexicality, Frequency and Length. The results are shown in Table 3. There were significant effects of Blocks (RTs reducing across blocks) and Lexicality (faster RTs to high frequency words than nonwords). There were also significant interactions involving Lexicality x Length and Blocks x Lexicality x Length. Lexicality effects were larger for long than short items, with the difference in RTs to long and short items reducing across blocks on day 1.

Model 3. Effects of Lexicality, Frequency, Blocks and Length on day 28

Model 3 analysed the data for day 28 only. The results are shown in Table 3. There was a significant effect of Lexicality and a significant Lexicality x Length interaction, with the lexicality effect again being greater for long than short items. Unlike Model 2, the impact of Blocks was not significant. Figure 1 shows that apart from some reduction of RTs for long nonwords between blocks 1 and 2, RTs changed little across blocks on day 28.

Discussion

RTs to English words and nonwords declined across blocks on day 1 as a result of repetition priming / word learning. There was good retention of the savings in RTs across the retention interval, resulting in significantly faster RTs on day 28 than day 1. The results for nonwords on day 1 were similar to those obtained by Maloney et al. (2009) and Kwok and Ellis (2014, 2015): RTs reduced decreased across the early blocks of trials, with the difference in RTs to short and long nonwords also reducing. Following Maloney et al. (2009) and Kwok and Ellis (2014, 2015) we would interpret that as evidence of a shift from sublexical to lexical reading as representations are created for the novel letter strings. If anything, RTs to long and short nonwords were slower to converge in Experiment 1 than in those previous studies, a point we will return to in the General Discussion. Words and nonwords showed good retention of day 1 savings across the 28-day retention interval. Using a different test of word recognition (thresholds rather than naming latencies), Salasoo et al. (1985) demonstrated retention of nonword learning over a 12-month interval following repeated exposure to nonwords in ten sessions spread over 12 days. The results for the nonwords on day 28 are similar to the results of Kwok and Ellis (2014, 205) with a slight slowing of RTs to the long nonwords in the first block following the retention interval and fast responses with small length effects thereafter.

The lexicality effect was greater for long than short stimuli on both days. On day 1, the lexicality effect was particularly marked in the early blocks when the nonwords were completely unfamiliar, reducing across blocks as RTs to the long nonwords in particular became faster as a result of learning.

None of the effects involving word frequency was significant, including the difference in length effects for high and low frequency words reported by Balota, Cortese, Sergent-Marshall, Spieler and Yap (2004), Weekes (1997) and Yap and Balota (2009). Possible reasons for the lack of frequency effects in Experiment 1 will be considered in the General Discussion along with other aspects of the results.

EXPERIMENT 2: REPEATED PRESENTATION OF HIGH FREQUENCY WORDS, LOW FREQUENCY WORDS AND NONWORDS ACROSS 28 DAYS IN SPANISH

Experiment 2 was modelled on Experiment 1 but with Spanish participants reading Spanish high frequency words, low frequency words and nonwords in ten blocks on day 1 and ten more blocks on day 28. We noted earlier the evidence for lexicality and frequency effects in Spanish, indicative of the use of lexical representations in a transparent orthography (e.g., Carreiras et al., 1993; Castejón et al., 2013; Cuetos & Suárez-Coalla, 2009; Davies et al., 2013; Perea & Carreiras, 1998). We therefore expected to see faster reading of high than low frequency words and faster reading of words than nonwords in Experiment 2. Effects of length on Spanish word naming have been reported by Cuetos and Barbón (2006) and Davies et al. (2013) but not investigated in detail. Possible interactions between lexicality and length or frequency and length have not been investigated in Spanish. The grain size theory of Ziegler et al. (2001) and Perry and Ziegler (2002) predicts that greater use of sublexical correspondences in a transparent orthography will result in larger length effects in Spanish compared with English word naming. Grain size theory also predicts slower convergence between RTs to short and long nonwords across repetitions in Spanish than in English because of the more persistent involvement of sublexical processing.

Method

Participants

The participants were 29 undergraduate students of the University of Oviedo, Spain (27 female, 2 male) with a mean age of 23.86 years (S.D. = 3.01; range 22 - 38). All were native speakers of Spanish with normal or corrected-to-normal vision and no history of reading or language problems. Participants received course credit in return for their participation. The experiment was approved by the Ethics Committee of the Department of Psychology, University of Oviedo.

Materials

As in Experiment 1, the experimental stimuli were 24 high frequency words, 24 low frequency words and 24 nonwords. Within each set, 12 items contained 4 letters while 12 contained 7 letters. Because of differences between the phonologies of Spanish and English, the 4-letter Spanish items contained two syllables rather than one while the 7-letter items contained three syllables rather than two. The short and long high frequency words, low frequency words and nonwords were matched on initial letters and phonemes. Twelve different onsets were used to make the items as distinct as possible. To optimise voice key activation, none of the stimuli began with a voiceless fricative.

Word frequencies were taken from SUBTLEX-ESP (Cuetos, Glez-Nosti, Barbon, & Brysbaert, 2011) which is based on words appearing in the subtitles of Spanish films and television programs. Word frequencies are expressed in Table 4 as Zipf values (van Heuven et al., 2014). The nonwords were matched to the real words on letter length, syllable length, initial letters and phonemes, and mean bigram frequency. The experimental words and nonwords are listed in Table 4 with their values on the matching variables. Eighteen additional high frequency words, low frequency words and nonwords (6 of each) were selected for use in practice trials.

(Table 4 about here)

Procedure

Participants were tested individually. As in Experiment 1, the participants were introduced to the task using 18 practice items. They then received 10 blocks of 72 stimuli (12 short high frequency words, 12 long high frequency words, 12 short low frequency words, 12 long low frequency words, 12 short nonwords and 12 long nonwords) presented in a different random order in each block. Words and nonwords were presented in black, lower case letters (System, 40 point, bold) on a white background. Each trial consisted of a centrally-presented fixation cross displayed for 1,000 ms, followed by the word or nonword stimulus for 2,000 ms, then a blank screen for 1,000 ms before the next trial began. At the end of each block, participants pressed the space bar when they were ready to start the next block. Participants wore headphones with a high-sensitivity microphone connected to a voice key that was linked to the computer. Presentation of the stimuli and recording of RTs was done using SuperLab Pro 2.0 (Cedrus Corporation, San Pedro, CA). The experimenter noted any trials in which the participant misread a nonword, hesitated or made a false start or other form of error. No feedback was given at any point.

Participants returned 28 days later for a second session which repeated the practice items and the 10 blocks of experimental stimuli using the same procedure as on day 1.

Results

Errors were removed and RTs trimmed as in Experiment 1. Naming errors, hesitations and failures to activate the voice key occurred on 618 trials (1.5% of the total). An additional 449 RTs were removed at the stage of RT trimming (1.1%), leaving 40,693 RTs for analysis (97.4% of the total). Table 5 (Supplemental materials) shows the accuracy and RT results for correct, trimmed responses in Experiment 2. Accuracy never fell below 91% correct for any stimulus on any block of trials and was at ceiling on day 28. For that reason, statistical analysis was again confined to the RT data.

Naming latencies (RTs)

Figure 2 shows the pattern of RTs for correct, trimmed responses to high frequency words, low frequency words and nonwords across days and blocks in Experiment 2. Log-transformed RTs were analysed using linear mixed effects modelling with three analyses mirroring models 1, 2 and 3 of Experiment 1. The data for the three analyses of Experiment 2 (Models 4 to 6) are presented in Appendix 2 and the R syntax in Appendix 9 (Supplemental materials).

(Figure 2 about here)

Model 4. Global analysis of RTs across days 1 and 28.

A global analysis of the Spanish data from days 1 and 28 was conducted to examine the effects of Day, Blocks, Lexicality, Frequency and Length defined as in Experiment 1. The results are shown in Table 3 with Experiment 2 analyses located alongside the corresponding analyses of Experiment 1 for ease of comparison.

There were significant effects of Day (faster RTs on day 28 than day 1) and Blocks (RTs decreasing across blocks over the two days) with a significant interaction between Day and Blocks (greater change in RTs across blocks on day 1 than day 28). The effect of Lexicality was significant (faster RTs to high frequency words than to nonwords) and there were significant interactions involving Lexicality x Length, Blocks x Lexicality x Length and Day x Lexicality x Length. Those interactions can be summarised by noting that the impact of length was greater for nonwords than high frequency words, with that difference decreasing across blocks and being greater on day 1 than day 28.

There were also significant Frequency x Length and Blocks x Frequency x Length interactions. The impact of length was greater for low than high frequency Spanish words, with that difference also decreasing across blocks.

Model 5. Effects of Lexicality, Frequency, Blocks and Length on day 1

Model 5 focused on the results for day 1, exploring the effects of Blocks, Lexicality, Frequency and Length. The results are shown in Table 3. There were significant effects of Blocks and Lexicality on day 1. As in the overall analysis, there were significant Frequency x Length and Blocks x Frequency x Length interactions, with the impact of length on day 1 being greater for low than high frequency words and decreasing across blocks. There were significant Lexicality x Length and Blocks x Lexicality x Length interactions, with the impact of length on day 1 being greater for nonwords than high frequency words and again decreasing across blocks.

Model 6. Effects of Lexicality, Frequency, Blocks and Length on day 28

Model 6 was the same as Model 5, but analysed the data from day 28 rather than day 1. The results are shown in Table 3. There were significant effects of Lexicality and Length on day 28 with interactions involving Lexicality x Length and Blocks x Lexicality x Length (a stronger effect of length on nonwords than high frequency words on day 28, with that effect reducing across blocks).

Discussion

The Spanish language results of Experiment 2 were similar to the English language results of Experiment 1 in several respects. As in Experiment 1, RTs declined significantly across blocks on day 1. There was good retention of the savings over the retention interval with the result that RTs were significantly faster on day 28. There is the suggestion of some improvement between blocks 1 and 2 of day 28 (Figure 2), but the effect of blocks on day 28 was not significant.

RTs were faster to Spanish high frequency words than to nonwords with the magnitude of the lexicality effect being similar to that for English in Experiment 1 (28 ms on day 1; 20 ms on day 28). The lexicality effect was again greater for long than for short stimuli. Unlike Experiment 1, the interaction between length, lexicality and blocks (stronger lexicality effect for long than for short items, particularly in early blocks of trials) was significant on day 28 as well as day 1. Frequency effects were apparent in Experiment 2 when they were not in Experiment 1. In particular, the frequency effect on day 1 was significantly greater for long than for short words, especially in the early blocks, a result similar to that seen for lexicality.

The evidence for lexicality, frequency and length effects in Spanish naming replicates the results of previous studies of reading in Spanish (e.g., Carreiras et al., 1993; Castejón et al., 2013; Cuetos & Barbón, 2006; Cuetos & Suárez-Coalla, 2009; Davies et al., 2013; Perea & Carreiras, 1998) and similar results reported for Italian (Colombo et al., 2006; Pagliuca et al., 2008; Paizi et al., 2010; Zoccolotti et al., 2009). The demonstration that lexicality and frequency effects in Spanish are stronger for long than short stimuli, especially for the initial presentations, is new. We are not aware of any previous studies of repetition priming of word recognition in Spanish, or of the maintenance of priming effects to words across a retention interval, but the results for Spanish were similar to those for English. Colombo et al. (2006) found that frequency effects reduced with repetition in Italian word naming.

The results for the Spanish nonwords in Experiment 2 look similar to those of Suárez-Coalla and Cuetos (2016) with the exception that RTs to short and long novel items appeared to converge more rapidly in the Suárez-Coalla and Cuetos (2016) than in the present Experiment 2. Possible reasons for this will be considered in the General Discussion.

COMBINED ANALYSIS OF THE RESULTS OF EXPERIMENT 1 (ENGLISH) AND EXPERIMENT 2 (SPANISH)

The remaining analyses compared RTs in English (Experiment 1) and Spanish (Experiment 2) directly. The emphasis was on length effects in the two languages, evaluating the predictions from grain size theory that sublexical letter-sound conversion contributes more to reading familiar words in transparent than opaque orthographies and that novel words make a faster and more complete transition from sublexical to lexical reading in opaque than transparent orthographies.

Figure 3 shows the effect of length on RTs to high frequency words, low frequency words and nonwords in English and Spanish across days 1 and 28. Length effects were measured as the difference in RTs to seven- and four-letter stimuli. Inspection of Figure 3 suggests that on day 1, length effects were present for both English and Spanish nonwords, decreasing across blocks but of similar magnitude. Length effects for words appear to have been greater for Spanish than for English, particularly for low frequency words. Figure 3 also suggests that length effects were greater for Spanish than for English on day 28 for both nonwords and words. Those indications were assessed in the analyses that follow.

(Figure 3 about here)

The first set of analyses concentrated on the results for day 1. Barr (2008) and Barr et al. (2013) recommended that analysis should be focused where effects are most apparent and changes greatest. For day 1 that was blocks one to six. Separate analyses were performed on RTs to high frequency words (model 7), low frequency words (model 8) and nonwords (model 9) in blocks one to six of day 1 with Language (English vs. Spanish), Blocks and Length as predictors (fixed effects). The results are shown in Table 6.

(Table 6 about here)

The indications of greater length effects in Spanish than in English on day 28 were assessed in a second set of analyses. Mirroring the analyses of day 1, separate analyses were performed on RTs to high frequency words (model 10), low frequency words (model 11) and nonwords (model 12) in blocks one to six of day 28 with Language, Blocks and Length as predictors (fixed effects). The results of those analyses are shown in Table 7.

The data for Models 7 to 12 are presented in Appendices 3 to 8 and the R syntax in Appendix 9 (Supplemental materials).

(Table 7 about here)

Model 7. Analysis of RTs in day 1, blocks 1 to 6 for high frequency words in English (Experiment 1) and Spanish (Experiment 2).

Model 7 analysed RTs to high frequency words from day 1, blocks one to six in the two experiments, exploring the effects of Language, Blocks and Length. There were significant effects of Language (faster RTs in English than Spanish) and Blocks (RTs becoming faster across blocks). A significant Blocks x Length interaction reflected a greater reduction in RTs across blocks for long than for short high frequency words.

Length effects were analysed for each language in each block ($\alpha = .008$; see Table 6). The mean length effect for English high frequency words on day 1 was 0 ms, with no significant effect of length in any of the six blocks. The mean length effect for Spanish high frequency words was 11 ms, reducing from 18 ms in block one to 4 ms in block six. The length effect was marginally significant in block one ($p = .009$) but not in blocks two to six.

Model 8. Analysis of RTs in day 1, blocks 1 to 6 for low frequency words in English and Spanish.

Model 8 analysed RTs to low frequency words from day 1, blocks 1 to 6 in the two experiments. There were significant effects of Language and Blocks, with significant interactions involving Blocks x Length (the length effect reducing across blocks) and, importantly, Language x Length.

The mean length effect for English low frequency words was 2 ms compared with 20 ms for Spanish. English low frequency words showed no significant effect of length in any of the six blocks. Length effects for Spanish low frequency words reduced from 46 ms in block one to 5 ms in block six and were significant in blocks one to three (Table 6).

Model 9. Analysis of RTs in day 1, blocks 1 to 6 for nonwords in English and Spanish.

Model 9 analysed RTs to nonwords from day 1, blocks one to six in the two experiments. There were significant effects of Language, Blocks and Length with a significant Blocks x Length interaction. The mean length effect for English nonwords was 40 ms, reducing from 84 ms in block one to 24 ms in block six and significant in blocks one to five. The mean length effect for Spanish nonwords was 44 ms, reducing from 78 ms in block one to 33 ms in block six and significant in all six blocks.

Model 10. Analysis of RTs in day 28, blocks 1 to 6 for high frequency words in English (Experiment 1) and Spanish (Experiment 2).

Model 10 analysed RTs to high frequency words from day 28, blocks one to six in the two experiments. The effect of Language was significant while the Language x Length interaction was approached significance (p = .088).

(Table 7 about here)

The mean length effect for English high frequency words on day 28 was -1 ms, with no significant effect of length in any of the six blocks. The mean effect for Spanish high frequency words was 16 ms, reducing from 27 ms in block one to 15 ms in block six and significant in blocks one to three (Table 7).

Model 11. Analysis of RTs in day 28, blocks 1 to 6 for low frequency words in English and Spanish.

Model 11 analysed RTs to low frequency words on day 28. There was a significant effect of Language and a significant Language x Length interaction. The mean length effect for English low frequency words was 0 ms, with no significant effect of length in any of the six blocks. Spanish low frequency words showed an overall length effect of 27 ms (from 46 ms in block one to 20 ms in block six) with significant length effects in all six blocks.

Model 12. Analysis of RTs in day 28, blocks 1 to 6 for nonwords in English and Spanish.

Model 12 analysed RTs to nonwords on day 28. The effects of Language and Length were significant and there was a significant Blocks x Length interaction. English nonwords showed an overall length effect of 15 ms. The length effect reduced from 46 ms in block one to -2 ms in block six. The effect was only significant in block one. Spanish nonwords showed an overall length effect of 31 ms, reducing from 60 ms in block one to 21 ms in block six, with a significant effect in blocks one to four.

Discussion

The main findings of the combined analyses can be summarised as follows.

High and low frequency English words showed no trace of a length effect on either day 1 or day 28. In contrast, length impacted on the naming of Spanish high frequency words in block one of day 1 (corrected p = .009) and blocks one to three of day 28. Spanish low frequency words showed significant length effects in blocks one to three of day 1 and in all six blocks on day 28. The length effects for Spanish words reduced across blocks on both days.

When nonwords were encountered for the first time in block one of day 1, the additional time per letter for seven- compared with four-letter nonwords was similar for English and Spanish (28 ms per additional letter for English and 26 ms for Spanish). This suggests that the ease or difficulty of applying sublexical letter-sound conversion to novel items is similar for skilled readers of English and Spanish. Length effects were significant for English nonwords in blocks one to five of day 1 and in block one on day 28. For Spanish nonwords, length effects were significant in blocks one to six of day 1 and blocks 1 to 4 of day 28. Figure 3 shows that length effects for Spanish low frequency words and nonwords were present throughout day 28 in a way that was not true for English low frequency words and nonwords.

General Discussion

The main focus of this paper is the possibility that English and Spanish differ in the relative use of lexical and sublexical conversion from orthography to phonology, as predicted by the grain size theory. Those differences should be detectable through comparisons of the impact of letter length on naming latencies in the two languages. We also sought to evaluate the possibility that as unfamiliar words become familiar through repeated exposure, skilled readers of English make a more complete switch from sublexical to lexical processing than occurs for skilled readers of Spanish. We will, however, begin this General Discussion section by commenting on some of the other similarities and differences between the results for the two languages.

Lexicality, frequency and the deployment of lexical representations

Lexicality effects were observed in both languages (as differences in naming RTs to high frequency words and nonwords). Frequency effects (differences in RTs to high and low frequency words) were found in Spanish but not English. Like previous authors, we take the presence of lexicality effects (and frequency effects in Spanish) to indicate that lexical representations are formed by readers of both transparent and opaque orthographies and are involved in converting familiar words from print to sound.

The differences in RTs to high and low frequency English words were small (8 ms on day 1 and 6 ms on day 28; see Table 2, supplemental materials) and non-significant. In Spanish, frequency effects were again small (11 ms on day 1; 8 ms on day 28; see Table 5, supplemental materials) but were significant on day 1. There are a number of reasons why the frequency effects may have been relatively small in the present experiments. First, frequency effects are generally smaller in word naming than in lexical decision where they are often studied (e.g., Balota & Spieler, 1999; Coane & Balota, 2010; see also Colombo et al., 2006, for similar results in Italian). Second, the high and low frequency words in Experiments 1 and 2 were interleaved with each other and with nonwords. When participants are presented in a reaction time task with stimuli containing an unpredictable mixture of relatively easy items and more difficult items, responses to the easier items tend to become slower while responses to the more difficult items become faster. Lupker, Brown, and Colombo (1997) found that when high and low frequency English words were interleaved, naming latencies to the low frequency words tended to become faster while latencies to the high frequency words became slower (compared with conditions in which the high and low frequency words were presented in separate blocks of trials. Using lexical decision rather than word naming, Rastle et al. (2003) reported a similar homogenisation of RTs to high and low frequency words in mixed blocks compared with 'pure' blocks. Pagliuca et al. (2008) found that when Italian high frequency words, low frequency words and nonwords were mixed together in a naming task, the frequency effect was reduced compared with a condition in which the stimuli were presented in pure blocks. Hence, the use of mixed blocks may have reduced the impact of word frequency in the present experiments.

Lexicality effects were similar in magnitude for English and Spanish and significant across both days in both languages. The lexicality effect for in block one of day 1 in English was 69 ms, reducing to an average of 23 ms across the last five blocks of day 1 and 16 ms on day 28. The lexicality effect for in block one of day 1 in Spanish was 55 ms, reducing to 22 ms across the last five blocks of day 1 and 20 ms on day 28. In both languages, the lexicality effect was stronger for long than short items, especially in earlier blocks. While the effects of lexicality were robust in both languages, they may still have been affected by interleaving. Rastle et al. (2003) found a reduction in the lexicality effect for English when high frequency words and nonwords were presented in mixed rather than pure blocks. Pagliuca et al. (2008) reported a similar result for Italian. We would expect both lexicality and frequency effects to be larger if pure blocks of stimuli were used rather than random interleaving.

Word learning and the durability of new lexical representations

Turning to the main focus of this paper, the two languages showed substantial length effects to nonwords when those items were presented for the first time. As noted above, we take this to be the hallmark of the involvement of sublexical letter-sound conversion in assembling the pronunciations of unfamiliar items encountered for the first time. The similarity of the length effects for nonwords encountered in block one of day 1 (84 ms in English, 78 ms in Spanish) suggests that there is no great difference between the languages in the ease with which skilled readers can assemble pronunciations for novel items using sublexical processes (fine grain mappings).

In both languages, RTs to the nonwords speeded up in the second and subsequent blocks on day 1. RTs to long nonwords reduced more than RTs to short nonwords, with the consequence that length effects reduced with repetition. We take this reflect the creation of larger-scale orthographic and phonological representations for the novel items, representations that are able to convert trained nonwords from written to spoken form on a more parallel, wholistic basis (cf. Coltheart et al., 2001; Kwok & Ellis, 2014, 2015; Maloney et al., 2009; Suárez-Coalla & Cuetos, 2016; Weekes, 1997).

There was good retention of savings in nonword RTs from day 1 to day 28 in both languages. Figures 1 and 2 suggest some loss of fidelity for the representations of long nonwords over the four-week retention interval, but one encounter with those items on day 28 was enough to restore RTs to the level they achieved at the end of the first training session. A similar phenomenon can be seen in the results of Kwok and Ellis (2014, 2015) and Suárez-Coalla and Cuetos (2016). Using a recognition threshold paradigm rather than naming, Salassoo et al. (1985) found evidence for retention of representations for newly-learned nonwords over a period of 12 months despite those items not having been experienced over the intervening months. Once lexical representations are formed for newly-learned items, they can survive for considerable periods without re-activation or rehearsal.

One difference between the results of Kwok and Ellis (2014, 2015) and those of the present Experiment 1 concerns the speed of the convergence between RTs to long and short nonwords across repetitions. Kwok and Ellis (2014, 2015) found that convergence for long and short English nonwords was virtually complete by block six or seven of day 1 but in the present Experiment 1, full convergence was only really apparent on day 28 (see Figures 1 and 3). Suárez-Coalla and Cuetos (2016) reported convergence of RTs for short and long unfamiliar words in Spanish at around block five (for items given no advance training) whereas in Experiment 2, length effects were present for Spanish nonwords on both days (Figures 2 and 3). There were procedural differences between the present Experiments 1 and 2 and the studies of Kwok and Ellis (2014, 2015) and Suárez-Coalla and Cuetos (2016) that may account for slower word learning in the present experiments. For example, Suárez-Coalla and Cuetos (2016) employed fewer stimuli than in the present Experiment 2, so there were fewer items to be learned, which may be a factor. Kwok and Ellis (2014, 2015) employed 12 short and 12 long nonwords, the same number as in the present experiments, so variation in the number of novel items to be learned cannot explain differences in the results of the present Experiment 1 compared with Kwok and Ellis (2014, 2015). The nonwords in the present experiments were, however, interleaved with familiar words in a way that was not true for Kwok and Ellis (2014, 2015) or Suárez-Coalla and Cuetos (2016). This meant that the intervals between successive presentations of the novel items in the present experiments were longer than in those previous studies, which may also influence word learning.

The presence of interleaved words in the present experiments also introduces the possibility of interference between familiar words and embryonic representations of novel items. We are not aware of studies that directly address the possibility of interference between novel words and interleaved familiar words, but learning of novel words has been shown to affect the processing of familiar words that resemble them, possibly as a result of the creation of new lexical neighbours (Bowers, Davis, & Hanley, 2005; Henderson, Weighall, Brown, & Gaskell, 2013; Lindsay & Gaskell, 2013). Learning new words may be particularly sensitive to the presence of interleaved familiar words that are similar in their orthography and phonology to the novel items. Further work is needed to clarify the factors that facilitate or inhibit word learning, but the numbers of items to be learned, the spacing of repetitions and the presence and similarity of interleaved familiar words are factors that seem to us to warrant further investigation.

Mechanisms of nonword and word learning

Conventionally, reduction in RTs to nonwords as a result of repetition and convergence of RTs to long and short items would be ascribed to the creation of lexical representations and a switch from sublexical to lexical reading. In contrast, the reduction across blocks in RTs to familiar words would be ascribed to repetition priming. We are struck, however, by the similarity in the learning functions for nonwords and words in the present experiments (cf. Grant & Logan, 1993; Lee, 1999). Dual-route models of reading lack learning mechanisms, so it is not possible to directly simulate either the creation of new lexical representations through learning or the strengthening of existing representations through repetition. Similarities between the functions generated by progressive switching from sublexical to lexical processing versus strengthening existing representations would presumably be largely coincidental.

There is, however, another way of thinking about the mechanisms responsible for learning novel words and strengthening the representations of familiar words through repetition. In distributed memory ('connectionist') models of word recognition, the processing units are letters (orthography) and phonemes (phonology). Familiar words are entrenched patterns of activation expressed across those processing units (e.g., Harm & Seidenberg, 1999, 2004; Plaut et al., 1996; Monaghan & Ellis, 2010). By definition, novel words and nonwords do not exist as established patterns in the network, but their component letter and phoneme sequences are likely to have been bound together by virtue of having been encountered in other, familiar words. For example, the letter and phoneme sequences contained in the English nonword *dast* will already have been encountered in words like *dark*, *dank*, *dust*, *last*, etc. Learning *dast* as a new word would require that connections between those component sequences should be strengthened to create a coherent processing unit. Importantly, that process of strengthening connections would be the same as the process that would strengthen the connections between the component letters and phonemes of *dank* if, for some reason, that low frequency word was to become more familiar through increased exposure. Similarities between the functions observed for learning novel words and repeating familiar words are to be expected within connectionist models because the mechanisms underlying the two forms of learning are the same (Rueckl, 1990; Stark & McClelland, 2000). These observations are relatively speculative because while length effects have been central to the development of the dual-route framework, they have hardly been investigated within connectionist models. More detailed observations concerning connectionist accounts of length effects can be found in Kwok and Ellis (2015).

Grain theory and the observed differences between English and Spanish

The present experiments were conducted to test two predictions derived from grain size theory. The first was that reading words in a transparent orthography will make greater use of sublexical letter-sound correspondences than reading words in an opaque orthography. The second was that reading newly-learned words in a transparent orthography will continue to be read by a combination of lexical and sublexical processes at a time when reading newly-learned words in an opaque orthography has switched from sublexical to lexical processing.

The evidence relating to those predictions will be considered shortly, but we will first address an unexpected finding in the results of Experiments 1 and 2.

Inspection of Figures 1 and 2 shows that the RTs obtained for both words and nonwords in Spanish were longer than in English. The difference between English and Spanish RTs was of similar magnitude for words and nonwords and resulted in significant effects of Language when that was included as a predictor in the combined analyses of the two experiments (Tables 6 and 7). The English and Spanish words and nonwords were matched on letter length, but the characteristics of English and Spanish phonology meant that the four-letter Spanish stimuli contained two syllables while the four-letter English stimuli contained only one syllable. Similarly, the seven-letter Spanish stimuli contained three syllables while their English counterparts contained only two. Spanish word naming latencies have been shown to be sensitive to length (Cuetos & Barbón, 2006; Davies et al., 2013), though the tight coupling between orthographic and phonological length in Spanish makes it very difficult to separate the effects of orthographic and phonological length. If, as we shall suggest below, Spanish makes more use than English of sublexical mappings between orthography and phonology, that and the need to map four letters onto two syllables rather than one, and seven letters onto three syllables rather than two, may have contributed to the overall differences in Spanish and English RTs observed here.

The results of the present experiments support the prediction from grain size theory that length effects should be more apparent in Spanish than English word naming. English word naming showed no detectable effects of length for either high or low frequency words. In contrast, Spanish high frequency words showed a marginally significant effect of length in block one of day 1 and a significant effect in blocks one to three of day 28 (Figure 3 and Tables 6 and 7). This suggests that a combination of lexical and sublexical processing are employed in reading Spanish high frequency words when they are encountered for the first time in a testing session (and interleaved with low frequency words and nonwords, which may be important). Predominantly lexical processing of Spanish high frequency words was achieved more successfully in the later blocks of day 1 than day 28 where participants may have been more alert to the intermixing of different types of stimuli.

Spanish low frequency words showed significant length effects in blocks one to three on day 1 and in all six blocks on day 28. The length effect for Spanish low frequency words encountered for the first time was larger (46 ms) than the comparable effect for Spanish high frequency words (18 ms). The length effect in block 1 of day 28 was also larger for Spanish low frequency words (46 ms) than for Spanish high frequency words (27 ms). These results suggest greater involvement of sublexical processing for low than high frequency Spanish words. On day 1 in particular, repeated presentation of the low frequency words appeared to boost the involvement of lexical processing with a consequent reduction in length effects in later blocks. Overall, the prediction from grain size theory that length effects will be more apparent in Spanish than English word naming is clearly supported by the present results.

Grain size theory also predicts that while lexical representations will be created for novel words (nonwords) in both transparent and opaque orthographies as a result of repeated encounters, the transition from sublexical to lexical processing will occur faster and more completely in opaque than transparent orthographies. That prediction was also supported here. We have noted that convergence of RTs to long and short English nonwords occurred more slowly in the present Experiment 1 than in the results of Kwok and Ellis (2014, 2015) and have suggested that increased spacing of repetitions and/or competition from interleaved familiar words may have contributed to that slower convergence. Nevertheless, RTs to long and short English nonwords converged substantially, particularly on day 28 when the length effect for English nonwords was only significant in block one and the effect across blocks six to ten was a mere 3 ms. In contrast, length effects for Spanish nonwords were substantial on both days (though only marginally significant in blocks five and six of day 28; see Table 7 and Figure 3).

Reading is not a natural, biologically-given skill but is acquired painstakingly as the result of education and experience. It is reasonable to expect that the structure and operation of a developing reading system will be influenced by the characteristics of the orthography being learned. We propose that the unreliability of sublexical letter-sound correspondences in English leads to the development of a reading system in which lexical mappings between orthography and phonology are favoured over sublexical mappings and word learning involves a relatively complete switch from sublexical to lexical processing. The transparency of Spanish orthography means that conflicts do not occur between sublexical and lexical processing. Mapping at all levels continues to be employed, though lexical mappings play an increasing role as words become more familiar. Reading of less familiar words and newly-learned words employs a combination of lexical and sublexical mappings at a point were the reading of similar items in English has moved to predominantly lexical processing.

Funding. This study was funded by Grant PSI2012-31913, from the Spanish Ministry of Economy and Competitiveness.

References

Alvarez, C. J., Carreiras, M., & Taft, M. (2001). Syllables and morphemes: Contrasting frequency effects in Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(2), 545 - 555. doi: 10.1037//0278-7393.27.2.545

Baayen, R. H. (2009). Corpus linguistics in morphology: morphological productivity. *Corpus linguistics. An international handbook*. Mouton De Gruyter, Berlin.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. doi: 10.1016/j.jml.2007.12.005

Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (release 2)[CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX lexical database [CD-ROM]: Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Balota, D., Cortese, M., Sergent-Marshall, S., Spieler, D., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*(2), 283-316. doi: 10.1037/0096-3445.133.2.283

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128*(1), 32-55. doi: 10.1037//0096-3445.128.1.32

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457-474. doi: 10.1016/j.jml.2007.09.002

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278. doi:10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. http://cran.R-project.org/package=lme4. R package version 0.999375-42.

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: the impact of novel word learning on the identification of visually similar words. *Cognition, 97,* 45- 54. doi: 10.1016/j.cognition.2005.02.002

Brysbaert, M., & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *The Quarterly Journal of Experimental Psychology, 64*(3), 545-559. doi: 10.1080/17470218.2010.503374

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977-990. doi: 10.3758/BRM.41.4.977

Carreiras, M., Alvarez, C. J., & Devega, M. (1993). Syllable frequency and visual word recognition in Spanish. *Journal of Memory and Language, 32*(6), 766-780. doi: 10.1006/jmla.1993.1038

Carreiras, M., Vergara, M., & Barber, H. (2005). Early event-related potential effects of syllabic processing during visual word recognition. *Journal of Cognitive Neuroscience, 17*(11), 1803-1817. doi: 10.1162/089892905774589217

Castejon, L., Rodriguez-Ferreiro, J., & Cuetos, F. (2013). Flexibilidad en el uso de estrategias de lectura de palabras en aprendices espanoles. [Flexibility in the use of word reading strategies in Spanish learners]. *Infancia y Aprendizaje, 36*(1), 51-60. doi: 10.1174/021037013804826564

Cave, C. B. (1997). Very long-lasting priming in picture naming. *Psychological Science, 8*(4), 322-325. doi: 10.1111/j.1467-9280.1997.tb00446.x

Coane, J. H., & Balota, D. A. (2010). Repetition priming across distinct contexts: Effects of lexical status, word frequency, and retrieval test. *The Quarterly Journal of Experimental Psychology, 63*(12), 2376-2398. doi: 10.1080/17470211003687546

Colombo, L., Pasini, M., & Balota, D. A. (2006). Dissociating the influence of familiarity and meaningfulness from word frequency in naming and lexical decision performance. *Memory & Cognition, 34*(6), 1312-1324. doi: 10.3758/BF03193274

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance, VI* (pp. 535-555). New York: Academic Press.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256. doi: 10.1037//0033-295X.108.1.204

Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *The Quarterly Journal of Experimental Psychology, 66*(5), 946-972. doi: 10.1080/17470218.2012.722660

Cuetos, F., & Barbón, A. (2006). Word naming in Spanish. *The European Journal of Cognitive Psychology, 18*, 415-436. doi: 10.1080/13594320500165896

Cuetos, F., & Suárez-Coalla, P. (2009). From grapheme to word in reading acquisition in Spanish. *Applied Psycholinguistics, 30*, 583-601. doi: 10.1017/S0142716409990038

Suárez-Coalla, P. & Cuetos, F. (2016). Semantic and phonological influences on visual word learning in a transparent orthography. *The Quarterly Journal of Experimental Psychology, in press.* doi: 10.1080/17470218.2016.1164733

Cuetos, F., Glez-Nosti, M., Barbón, A., & Brysbaert, M. (2011). SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicológica: Revista de Metodología y Psicología Experimental*, *32*(2), 133-143.

Davies, R., Barbón, A., & Cuetos, F. (2013). Lexical and semantic age-of-acquisition effects on word naming in Spanish. *Memory & Cognition, 41*(2), 297-311. doi: 10.3758/s13421-012-0263-8

Davies, R., Rodríguez-Ferreiro, J., Suárez, P., & Cuetos, F. (2013). Lexical and sub-lexical effects on accuracy, reaction time and response duration: impaired and typical word and pseudoword reading in a transparent orthography. *Reading and Writing*, *26*(5), 1-18. doi: 10.1007/s11145-012-9388-1

Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods, 37*(4), 665-671. doi: 10.3758/BF03192738

Duyck, W., Desmet, T., Verbeke, L. P., & Brysbaert, M. (2004). WordGen: A tool for word selection and nonword generation in Dutch, English, German, and French. *Behavior Research Methods, Instruments, & Computers, 36*(3), 488-499. doi: 10.3758/BF03195595

Ellis, A. W., Ferreira, R., Cathles-Hagan, P., Holt, K., Jarvis, L., & Barca, L. (2009). Word learning and the cerebral hemispheres: from serial to parallel processing of written words. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1536), 3675-3696. doi: 10.1098/rstb.2009.0187

González-Nosti, M., Barbón, A., Rodríguez-Ferreiro, J., & Cuetos, F. (2013). Effects of the psycholinguistic variables on the lexical decision task in Spanish: A study with 2,765 words. *Behavior Research Methods*, *46*(2), 1-9. doi: 10.3758/s13428-013-0383-5

Grant, S. C., & Logan, G. D. (1993). The loss of repetition priming and automaticity over time as a function of degree of initial learning. *Memory & Cognition*, *21*(5), 611-618. doi: 10.3758/BF03197193

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychological Review*, *106*(3), 491-528. doi: 10.1037/0033-295X.106.3.491

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*(3), 662-720. doi: 10.1037/0033-295X.111.3.662

Henderson, L., Weighall, A., Brown, H., & Gaskell, G. (2013). Online lexical competition during spoken word recognition and word learning in children and adults. *Child Development*, *84*(5), 1668-1685. doi: 10.1111/cdev.12067

Hogaboam, T., & Perfetti, C. (1978). Reading skill and the role of verbal experience in decoding. *Journal of Educational Psychology, 70*(5), 717-729. doi: 10.1037//0022-0663.70.5.717

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *lmerTest: tests in linear mixed effects models*. R package version 2.0-20.

Kwok, R. K. W., & Ellis, A. W. (2014). Visual word learning in adults with dyslexia. *Frontiers in Human Neuroscience*, *8*, 1-12. doi: 10.3389/fnhum.2014.00264

Kwok, R. K. W., & Ellis, A. W. (2015). Visual word learning in skilled readers of English. *The Quarterly Journal of Experimental Psychology*, *68*(2), 326-349. doi: 10.1080/17470218.2014.944549

Lee, C. H. (1999). A locus of the word-length effect on word recognition. *Journal of Reading Psychology, 20*(2)*,* 129-150. doi: 10.1080/027027199278448.

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 608-622. doi: 10.1037/a0029243

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(3), 570-590. doi: 10.1037//0278-7393.23.3.570

Maloney, E., Risko, E. F., O'Malley, S., & Besner, D. (2009). Tracking the transition from sublexical to lexical processing: On the creation of orthographic and phonological lexical representations. *The Quarterly Journal of Experimental Psychology, 62*(5), 858-867. doi: 10.1080/17470210802578385

Mason, M. (1978). From print to sound in mature readers as a function of reader ability and two forms of orthographic regularity. *Memory & Cognition, 6*(5), 568-581. doi: 10.3758/BF03198246

Monaghan, P., & Ellis, A. W. (2010). Modeling reading development: Cumulative, incremental learning in a computational model of word naming. *Journal of Memory and Language, 63*(4), 506-525. doi: 10.1016/j.jml.2010.08.003

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers, 36*(3), 516-524. doi: 10.3758/BF03195598

Pagliuca, G., Arduino, L. S., Barca, L. & Burani, C. (2008). Fully transparent orthography, yet lexical reading aloud: The lexicality effect in Italian. *Language & Cognitive Processes, 23*(3), 422-433. doi: 10.1080/01690960701626036

Paizi, D., Burani, C., & Zoccolotti, P. (2010). List context effects in reading Italian words and nonwords: Can the word frequency effect be eliminated? *European Journal of Cognitive Psychology, 22*(7), 1039-1065. doi:10.1080/09541440903216492

Perea, M., & Carreiras, M. (1998). Effects of syllable frequency and syllable neighborhood frequency in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 24*(1), 134-144. doi: 10.1037//0096-1523.24.1.134

Perea, M., Carreiras, M., & Grainger, J. (2004). Blocking by word frequency and neighborhood density in visual word recognition: A task-specific response criteria account. *Memory & Cognition, 32*(7), 1090-1102. doi:10.3758/BF03196884

Perry, C., & Ziegler, J. C. (2002). Cross-language computational investigation of the length effect in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 28*(4), 990-1001. doi: 10.1037//0096-1523.28.4.990

Pinheiro, J., & Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56-115. doi:10.1037/0033-294X.103.1.56.

Rastle, K., & Coltheart, M. (1994). Serial processes in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1197-1211. doi: 10.1037/0096.1523.20.6.1197

Rastle, K., Kinoshita, S., Lupker, S. J., & Coltheart, M. (2003). Cross-task strategic effects. *Memory & Cognition, 31*(6), 867-876. doi: 10.3758/BF03196441

Rueckl, J. G. (1990). Similarity effects in word and pseudoword repetition priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 374-391. doi: 10.1037/0278-7393.16.3.374.

Salasoo, A., Shiffrin, R. M., & Feustel, T. C. (1985). Building permanent memory codes: Codification and repetition effects in word identification. *Journal of Experimental Psychology: General, 114*(1), 50-77. doi: 10.1037//0096-3445.114.1.50

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3*(1), 1-17. doi: 10.1037//0096-1523.3.1.1

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*: Psychology Software Incorporated.

Sebastián-Gallés, N., Martí, M. A, Carreiras, M., & Cuetos, F. (2000). *Lexesp, Léxico informatizado del Español*. Barcelona: Universidad de Barcelona.

Stark, C. E. L., & McClelland, J. L. (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*(4), 945-972. doi: 10.1037/0278-7393.26.4.945

Suárez-Coalla, P., Álvarez-Cañizo, M., & Cuetos, F. (2014). Orthographic learning in Spanish children. *Journal of Research in Reading*, *2*, 166-181. doi: 10.1111/1467-9817.12043

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology, 67*(3), 1176-1190. doi: 10.1080/17470218.2013.850521

Weekes, B. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology, 50* (2), 439-456. doi: 10.1080/027249897392170

Wiggs, C. L., Weisberg, J., & Martin, A. (2006). Repetition priming across the adult lifespan—the long and short of it. *Aging, Neuropsychology, and Cognition, 13*(3), 308-325. doi: 10.1080/138255890968718

Yap, M., & Balota, D. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language, 60*(4), 502-529. doi: 10.1016/j.jml.2009.02.001

Ziegler, J., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic Grain Size Theory. *Psychological Bulletin, 131*(1), 3-29. doi: 10.1037/0033-2909.131.1.3

Ziegler, J., Perry, C., & Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: An extension to German. *European Journal of Cognitive Psychology, 12*(3), 413-430. doi: 10.1080/09541440050114570

Ziegler, J., Perry, C., Jacobs, A., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science, 2*(5), 379-384. doi: 10.1111/1467-9280.00370

Zoccolotti, P., De Luca, M., Di Filipp, G., Judica, A., & Martelli, M. (2009). Reading development in an orthographically regular language: effects of length, frequency, lexicality and global processing ability. *Reading and Writing, 93*(9)*,* 1053-1079. doi:10.1007/s11145-008-9144-8

Figure legends

Figure 1. *Mean naming RTs across blocks and testing sessions for high frequency words, low frequency words and nonwords in Experiment 1 (English).*

Figure 2. *Mean naming RTs across blocks and testing sessions for high frequency words, low frequency words and nonwords in Experiment 2 (Spanish).*

Figure 3. *Length effects for high frequency words, low frequency words and nonwords in Experiments 1 (English) and 2 (Spanish).*

Table 1. Details of the stimuli used in Experiment 1 (English). Word frequencies are presented as Zipf values (log10 of the frequency per billion words of text; van Heuven et al., 2014).

	4 letters			7 letters		
	High frequency	Low frequency	Nonwords	High frequency	Low frequency	Nonwords
SUBTLEX word frequency (Zipf values)						
Mean	5.12	3.79	–	5.02	3.65	–
SD	0.20	0.31	–	0.24	0.47	–
Mean log bigram frequency						
Mean	3.38	3.32	3.38	3.38	3.30	3.39
SD	0.15	0.22	0.12	0.10	0.19	0.09

High frequency words. 4-letter: beat, card, cost, deal, king, mark, news, pick, poor, rest, team, wear; *7-letter:* believe, contract, country, darling, kitchen, machine, nervous, perhaps, promise, respect, teacher, welcome.

Low frequency words. 4-letter: bake, cart, cord, deed, kite, mute, nest, pier, plug, ripe, tart, wolf; *7-letter:* biscuit, concert, concise, default, ketchup, mermaid, neutral, perfume, profile, rubbish, toaster, wealthy.

Nonwords. 4-letter: blop, carg, cont, dast, kest, marb, nate, pite, plid, rell, tond, wost; *7-letter:* bencort, carklin, coftrip, dempton, kintore, marpoon, nelpoon, pembert, plinore, roffler, tismole, wedrick.

Table 3. Summary of linear mixed effects (LME) models of $\log_{10}$(RTs) naming latencies (RTs) in English (Experiment 1) and Spanish (Experiment 2), followed by a combined analysis of data from day 1, blocks 1 to 7 in both experiments. Where the same analyses were performed on the data from Experiments 1 and 2, they have been placed side-by-side in the table to facilitate comparison.

	Model 1 Experiment 1: English (Day 1 and 28) *Factors*: Day, Blocks, Frequency, Lexicality, Length				Model 4 Experiment 2: Spanish (Day 1 and 28) *Factors*: Day, Blocks, Frequency, Lexicality, Length			
	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*
Intercept	2.698	0.011	235.569	<.0001***	2.800	0.006	438.544	<.0001***
Day	-0.023	0.004	-6.277	<.0001***	-0.035	0.002	-16.893	<.0001***
Blocks	-0.003	0.001	-4.679	<.0001***	-0.002	0.001	-3.589	<.0001***
Frequency	0.006	0.007	0.955	.341	0.004	0.004	0.898	.371
Lexicality	0.015	0.007	2.313	.022*	0.014	0.004	3.262	< .001**
Length	0.003	0.007	0.457	.648	0.008	0.005	1.763	.080
Day x Blocks	0.004	0.001	6.073	<.0001***	0.003	<.001	7.665	<.0001***
Day x Frequency	0.001	0.005	0.125	.900	<.001	0.003	0.042	.967
Day x Lexicality	-0.003	0.005	-0.494	.621	-0.002	0.003	-0.627	.530
Day x Length	-0.002	0.005	-0.360	.719	0.005	0.003	1.734	.083
Blocks x Frequency	<.001	0.001	-0.127	.899	<.001	<.001	1.292	.197
Blocks x Lexicality	-0.001	0.001	-1.237	.216	<.001	<.001	-1.113	.266
Blocks x Length	-0.001	0.001	-1.031	.303	<.001	<.001	-0.395	.693
Frequency x Length	<.001	0.009	0.001	.999	0.013	0.006	2.130	.035*
Lexicality x Length	0.039	0.009	4.212	<.0001***	0.027	0.006	4.580	<.0001***
Day x Blocks x Frequency	<.001	0.001	-0.324	.746	-0.001	<.001	-1.084	.278
Day x Blocks x Lexicality	<.001	0.001	-0.070	.944	<.001	<.001	0.208	.835
Day x Blocks x Length	<.001	0.001	-0.176	.860	<.001	<.001	-0.913	.361

Day x Frequency x Length	0.003	0.007	0.481	.631	-0.003	0.004	-0.667	.505
Day x Lexicality x Length	-0.020	0.007	-2.792	<.001**	-0.012	0.004	-2.997	< .001**
Blocks x Frequency x Length	<.001	0.001	0.067	.946	-0.002	<.001	-3.692	<.0001***
Blocks x Lexicality x Length	-0.003	0.001	-3.517	<.0001***	-0.002	<.001	-4.659	<.0001***
Day x Blocks x Frequency x Length	-0.001	0.001	-0.623	.533	0.001	0.001	1.454	.146
Day x Blocks x Lexicality x Length	0.002	0.001	1.395	.163	0.001	0.001	1.313	.189
Variance components, random effects	*SD*				*SD*			
Item	0.013				0.009			
Subject	0.052				0.030			
Blocks	0.003				0.003			
Length	0.013				0.009			
Residual	0.065				0.039			
	Model 2 Experiment 1: English (Day 1 only) *Factors*: Blocks, Frequency, Lexicality, Length				Model 5 Experiment 2: Spanish (Day 1 only) *Factors*: Blocks, Frequency, Lexicality, Length			
	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*
Intercept	2.698	0.013	208.147	<.0001***	2.800	0.007	383.363	<.0001***
Blocks	-0.003	0.001	-4.028	<.0005***	-0.002	0.001	-3.401	< .001**
Frequency	0.006	0.007	0.905	.367	0.004	0.004	0.933	.353
Lexicality	0.015	0.007	2.209	.029*	0.014	0.004	3.496	<.0001***
Length	0.003	0.007	0.435	.664	0.008	0.005	1.768	.080
Blocks x Frequency	<.001	0.001	-0.117	.907	<.001	<.001	1.549	.121
Blocks x Lexicality	-0.001	0.001	-1.269	.204	<.001	<.001	-1.125	.261
Blocks x Length	-0.001	0.001	-1.065	.287	<.001	<.001	-0.450	.653
Frequency x Length	<.001	0.010	0.009	.992	0.013	0.005	2.345	.020*
Lexicality x Length	0.039	0.010	4.018	< .001***	0.027	0.005	5.008	<.0001***
Blocks x Frequency x Length	<.001	0.001	0.052	.959	-0.002	<.001	-4.204	<.0001***
Blocks x Lexicality x Length	-0.003	0.001	-3.618	<.0005***	-0.002	<.001	-5.302	<.0001***
Variance components, random effects	*SD*				*SD*			

Item (Intercept)	0.014				0.008			
Subject (Intercept)	0.060				0.036			
Blocks	0.004				0.003			
Length	0.015				0.013			
Residual	0.063				0.035			
	Model 3 Experiment 1: English (Day 28 only) *Factors*: Blocks, Frequency, Lexicality, Length				Model 6 Experiment 2: Spanish (Day 28 only) *Factors*: Blocks, Frequency, Lexicality, Length			
	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*
(Intercept)	2.675	0.011	245.155	<.0001***	2.765	0.007	419.816	<.0001***
Blocks	<.001	0.001	0.092	.927	<.001	0.001	0.601	.552
Frequency	0.007	0.006	1.110	.269	0.004	0.004	0.848	.398
Lexicality	0.013	0.006	2.025	.045*	0.012	0.004	2.644	< .001**
Length	0.001	0.007	0.209	.835	0.013	0.005	2.562	.011*
Blocks x Frequency	<.001	0.001	-0.606	.544	<.001	<.001	-0.179	.858
Blocks x Lexicality	-0.001	0.001	-1.370	.171	<.001	<.001	-0.795	.426
Blocks x Length	-0.001	0.001	-1.320	.187	-0.001	<.001	-1.798	.072
Frequency x Length	0.003	0.009	0.393	.695	0.010	0.006	1.585	.116
Lexicality x Length	0.019	0.009	2.145	.033*	0.015	0.006	2.374	.019*
Blocks x Frequency x Length	-0.001	0.001	-0.834	.404	-0.001	<.001	-1.730	.084
Blocks x Lexicality x Length	-0.001	0.001	-1.592	.112	-0.001	<.001	-2.929	< .001**
Variance components, random effects	*SD*				*SD*			
Item (Intercept)	0.013				0.010			
Subject (Intercept)	0.050				0.031			
Blocks	0.003				0.004			
Length	0.011				0.014			
Residual	0.063				0.038			

Note. *$p < .05$, **$p < .01$, ***$p < .001$.

Table 4. Details of the stimuli used in Experiment 2 (Spanish). Word frequencies are presented as Zipf values (log10 of the frequency per billion words of text; van Heuven et al., 2014).

	4 letters			7 letters		
	High frequency	Low frequency	Nonwords	High frequency	Low frequency	Nonwords
SUBTLEX-ESP word frequency (Zipf values)						
Mean	4.60	3.55	–	4.23	3.27	–
SD	0.46	0.57		0.39	0.54	–
Mean log bigram frequency						
Mean	2.03	2.04	2.04	1.85	1.85	1.85
SD	0.54	0.59	2.10	0.41	0.41	0.41

Note. *High frequency words, 4-letter:* baño, beso, dedo, doce, malo, mito, nota, paro, poca, rato, rico, tela; *7-letter:* batalla, belleza, derrota, dominio, maestro, milagro, notable, palacio, postura, radical, revista, terreno.

Low frequency words, 4-letter: bazo, bono, divo, dona, mago, mimo, neto, pato, popa, rana, rezo, teja; *7-letter:* bazofia, bofetón, derrame, docente, maligno, mitosis, nodriza, palanca, podrida, racismo, recital, tenedor.

Nonwords, 4-letter: bapo, beno, deco, doba, maco, mifo, nupa, pafo, poga, rada, rego, teba; *7-letter:* bagafia, betenza, derrupa, dotenio, malesno, mitegro, nograza, pabenco, pogriba, rabesal, redutal, temebor.

Table 6. Combined analysis of RT data from day 1, blocks 1 to 6 in Experiment 1 (English) and Experiment 2 (Spanish) using linear mixed effects (LME) models.

Model 7
Experiments 1 & 2: English and Spanish
Day 1, blocks 1 to 6, high frequency words only
Factors: Language, Blocks, Length

	Estimate	*SE*	*t*	*p*
Intercept	2.703	0.010	258.056	<.0001***
Language	0.100	0.014	6.953	<.0001***
Blocks	-0.005	0.001	-3.886	<.0001***
Length	0.009	0.007	1.309	.194
Language x Blocks	0.002	0.002	1.081	.283
Language x Length	0.002	0.010	0.192	.848
Blocks x Length	-0.003	0.001	-2.809	<.01**
Language x Blocks x Length	0.001	0.001	1.109	.267
Variance components, random effects		*SD*		
Item		0.013		
Subject		0.047		
Blocks		0.006		
Length		0.013		
Residual		0.046		

		English high frequency words		Spanish high frequency words	
		Estimate	*p*	Estimate	*p*
Block 1	Intercept	2.706		2.804	
	Length (α = .008)	0.007	.278	0.010	.009
Block 2	Intercept	2.686		2.795	
	Length	0.005	.471	0.006	.104
Block 3	Intercept	2.682		2.790	
	Length	0.002	.801	0.006	.105
Block 4	Intercept	2.683		2.787	
	Length	-0.003	.697	0.009	.022
Block 5	Intercept	2.678		2.787	
	Length	-0.006	.331	0.006	.108
Block 6	Intercept	2.672		2.787	
	Length	-0.003	.673	0.002	.488

Model 8
Experiments 1 & 2: English and Spanish
Day 1, blocks 1 to 6, low frequency words only
Factors: Language, Blocks, Length

	Estimate	*SE*	*t*	*p*
Intercept	2.714	0.011	249.103	<.0001***
Language	0.094	0.015	6.321	<.0001***
Blocks	-0.007	0.001	-4.639	<.0001***

Length	0.007	0.007	1.078	.284
Language x Blocks	0.003	0.002	1.663	.101
Language x Length	0.021	0.010	2.177	.032*
Blocks x Length	-0.002	0.001	-2.221	.026*
Language x Blocks x Length	-0.002	0.001	-1.761	.078

Variance components, random effects	*SD*
Item	0.012
Subject	0.050
Blocks	0.007
Length	0.014
Residual	0.047

		English low frequency words		Spanish low frequency words	
		Estimate	*p*	Estimate	*p*
Block 1	Intercept	2.715		2.812	
	Length (α = .008)	0.010	.139	0.026	<.0001***
Block 2	Intercept	2.692		2.798	
	Length	0.001	.953	0.016	<.0001***
Block 3	Intercept	2.690		2.793	
	Length	-0.005	.508	0.012	<.0001***
Block 4	Intercept	2.688		2.794	
	Length	-0.003	.645	0.008	.026
Block 5	Intercept	2.677		2.790	
	Length	-0.001	.965	0.009	.021
Block 6	Intercept	2.676		2.792	
	Length	-0.005	.461	0.003	.327

Model 9

Experiments 1 & 2: English and Spanish

Day 1, blocks 1 to 6, nonwords only

Factors: Language, Blocks, Length

	Estimate	*SE*	*t*	*p*
Intercept	2.721	0.012	222.490	<.0001***
Language	0.097	0.017	5.819	<.0001***
Blocks	-0.007	0.001	-5.071	<.0001***
Length	0.052	0.008	6.916	<.0001***
Language x Blocks	0.003	0.002	1.508	.136
Language x Length	-0.008	0.010	-0.805	.422
Blocks x Length	-0.006	0.001	-6.329	<.0001***
Language x Blocks x Length	0.001	0.001	1.013	.311

Variance components, random effects	*SD*
Item	0.012
Subject	0.057
Blocks	0.006
Length	0.021
Residual	0.052

	English nonwords		Spanish nonwords	
	Estimate	*p*	Estimate	*p*

Block 1	Intercept	2.728		2.820	
	Length (α = .008)	0.057	<.0001***	0.045	<.0001***
Block 2	Intercept	2.697		2.806	
	Length	0.028	<.0001***	0.032	<.0001***
Block 3	Intercept	2.685		2.802	
	Length	0.028	<.0001***	0.024	<.0001***
Block 4	Intercept	2.689		2.799	
	Length	0.0215	.003*	0.018	<.0001***
Block 5	Intercept	2.683		2.800	
	Length	0.025	.001**	0.016	<.0001***
Block 6	Intercept	2.684		2.794	
	Length	0.017	.024	0.020	<.0001***

Note. *p < .05, **p < .01, ***p < .001 for main analyses. Corrected thresholds for pairwise comparisons of length effects in blocks 1 to 6 (p / 6): *α < .008, **α < .002, ***α < .0002.

Table 7. Combined analysis of RT data from day 28, blocks 1 to 6 in Experiment 1 (English) and Experiment 2 (Spanish) using linear mixed effects (LME) models.

Model 10
Experiments 1 & 2: English and Spanish
Day 28, blocks 1 to 6, high frequency words only
Factors: Language, Blocks, Length

	Estimate	*SE*	*t*	*p*
Intercept	2.679	0.009	285.062	<.0001***
Language	0.087	0.013	6.764	<.0001***
Blocks	-0.001	0.001	-0.974	.333
Length	0.001	0.007	0.140	.889
Language x Blocks	0.001	0.002	0.629	.532
Language x Length	0.016	0.009	1.722	.088
Blocks x Length	-0.001	0.001	-0.615	.539
Language x Blocks x Length	-0.001	0.001	-0.898	.369

Variance components, random effects	*SD*
Item	0.012
Subject	0.042
Blocks	0.006
Length	0.013
Residual	0.048

		English high frequency words		Spanish high frequency words	
		Estimate	*p*	Estimate	*p*
Block 1	Intercept	2.687		2.771	
	Length (α = .008)	0.002	.790	0.018	<.0001***
Block 2	Intercept	2.668		2.763	
	Length	0.005	.469	0.010	.006*
Block 3	Intercept	2.669		2.759	
	Length	0.001	.916	0.011	.002**
Block 4	Intercept	2.675		2.766	
	Length	0.007	.270	0.007	.040
Block 5	Intercept	2.675		2.767	
	Length	0.004	.509	0.006	.077
Block 6	Intercept	2.673		2.766	
	Length	0.001	.852	0.010	.010

Model 11
Experiments 1 & 2: English and Spanish
Day 28, blocks 1 to 6, low frequency words only
Factors: Language, Blocks, Length

	Estimate	*SE*	*t*	*p*
Intercept	2.680	0.009	294.942	<.0001***
Language	0.092	0.012	7.398	<.0001***
Blocks	0.001	0.001	0.388	.699

Length		0.004	0.007	0.622	.534
Language x Blocks		-0.001	0.002	-0.750	.456
Language x Length		0.024	0.010	2.484	.014*
Blocks x Length		-0.001	0.001	-1.222	.222
Language x Blocks x Length		-0.002	0.001	-1.235	.217
Variance components, random effects			*SD*		
Item			0.012		
Subject			0.040		
Blocks			0.006		
Length			0.014		
Residual			0.050		
		English low frequency words		Spanish low frequency words	
		Estimate	*p*	Estimate	*p*
Block 1	Intercept	2.687		2.777	
	Length (α = .008)	0.001	.896	0.029	<.0001***
Block 2	Intercept	2.675		2.765	
	Length	0.004	.503	0.020	<.0001***
Block 3	Intercept	2.679		2.766	
	Length	0.001	.999	0.015	<.0001***
Block 4	Intercept	2.681		2.767	
	Length	0.005	.424	0.017	<.0001***
Block 5	Intercept	2.674		2.769	
	Length	0.001	.928	0.012	<.0001***
Block 6	Intercept	2.691		2.798	
	Length	0.007	.288	0.014	<.0001***
Model 12 Experiments 1 & 2: English and Spanish Day 28, blocks 1 to 6, nonwords only *Factors*: Language, Blocks, Length					
		Estimate	*SE*	*t*	*p*
Intercept		2.688	0.010	278.111	<.0001***
Language		0.090	0.013	6.763	<.0001***
Blocks		-0.001	0.001	-0.782	.437
Length		0.028	0.007	4.051	<.0001***
Language x Blocks		0.001	0.002	0.394	.695
Language x Length		0.007	0.010	0.753	.453
Blocks x Length		-0.004	0.001	-4.436	<.0001***
Language x Blocks x Length		<.001	0.001	0.236	.814
Variance components, random effects			*SD*		
Item			0.012		
Subject			0.043		
Blocks			0.006		
Length			0.015		
Residual			0.051		
		English nonwords		Spanish nonwords	
		Estimate	*p*	Estimate	*p*

Block 1	Intercept	2.683		2.783	
	Length (α = .008)	0.034	<.0001***	0.037	<.0001***
Block 2	Intercept	2.682		2.775	
	Length	0.010	.166	0.020	<.0001***
Block 3	Intercept	2.685		2.770	
	Length	0.008	.238	0.021	<.0001***
Block 4	Intercept	2.682		2.777	
	Length	0.009	.195	0.017	<.0001***
Block 5	Intercept	2.678		2.777	
	Length	0.014	.050	0.012	.014
Block 6	Intercept	2.689		2.777	
	Length	0.001	.997	0.014	.016

Note. *p < .05, **p < .01, ***p < .001 for main analyses. Corrected thresholds for pairwise comparisons of length effects in blocks 1 to 6 (p / 6): *α < .008, **α < .002, ***α < .0002.

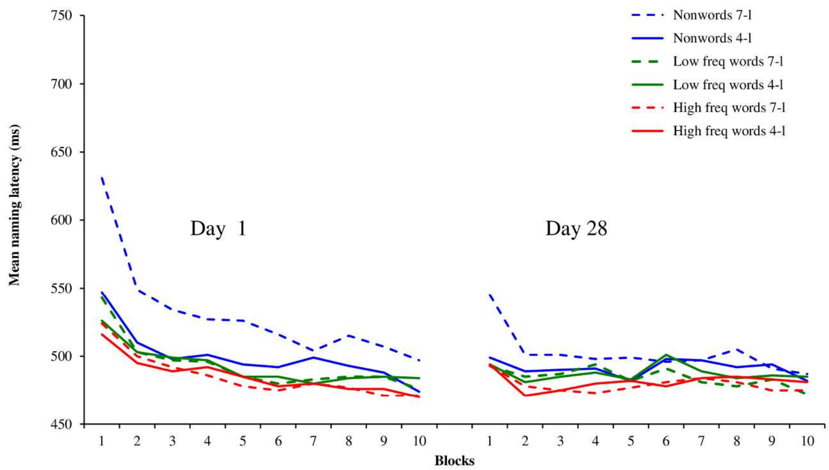

Figure 1. Mean naming RTs across blocks and testing sessions for high frequency words, low frequency words and nonwords in Experiment 1 (English).

297x210mm (200 x 200 DPI)

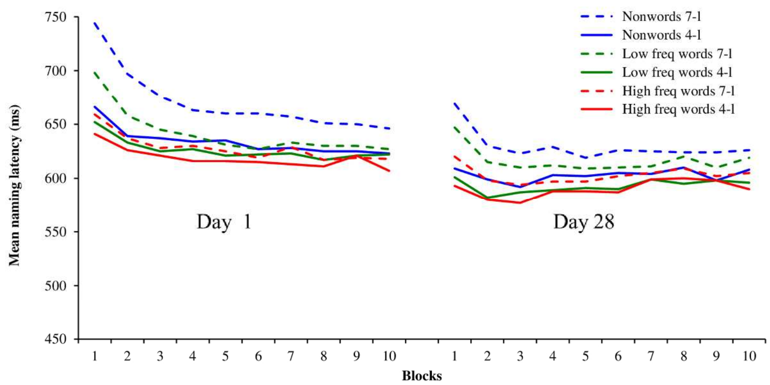

Figure 2. Mean naming RTs across blocks and testing sessions for high frequency words, low frequency words and nonwords in Experiment 2 (Spanish).

297x210mm (200 x 200 DPI)

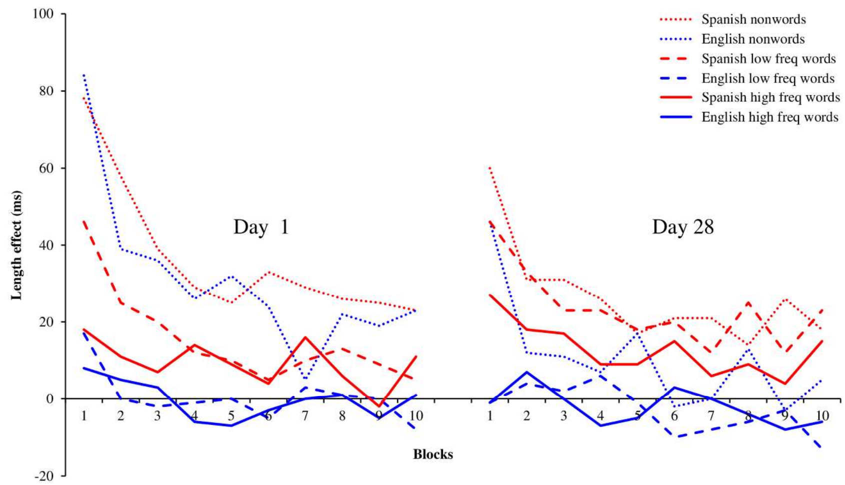

Figure 3. Length effects for high frequency words, low frequency words and nonwords in Experiments 1 (English) and 2 (Spanish).

297x210mm (200 x 200 DPI)

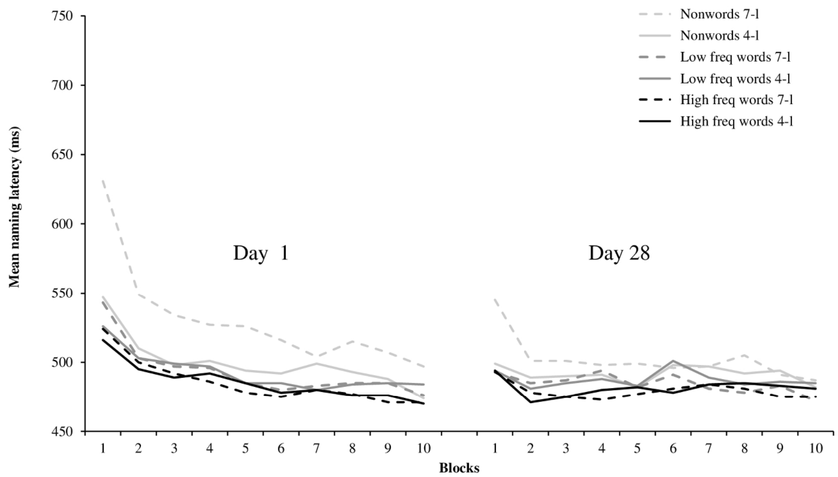

Figure 1. Mean naming RTs across blocks and testing sessions for high frequency words, low frequency words and nonwords in Experiment 1 (English).

297x210mm (200 x 200 DPI)

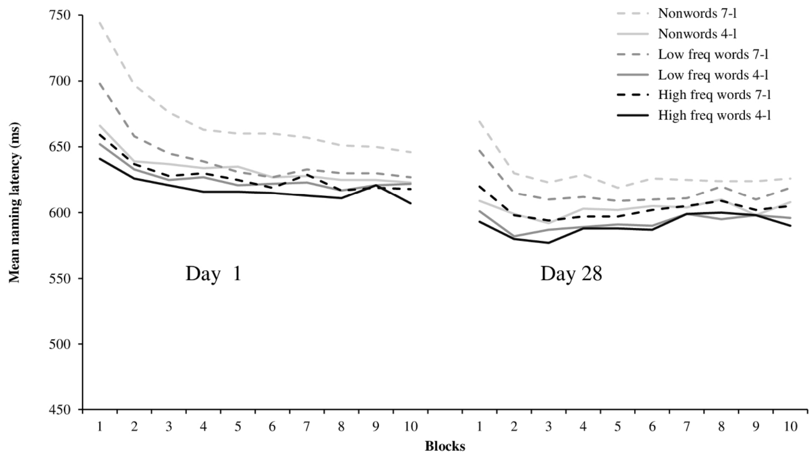

Figure 2. Mean naming RTs across blocks and testing sessions for high frequency words, low frequency words and nonwords in Experiment 2 (Spanish).

297x210mm (200 x 200 DPI)

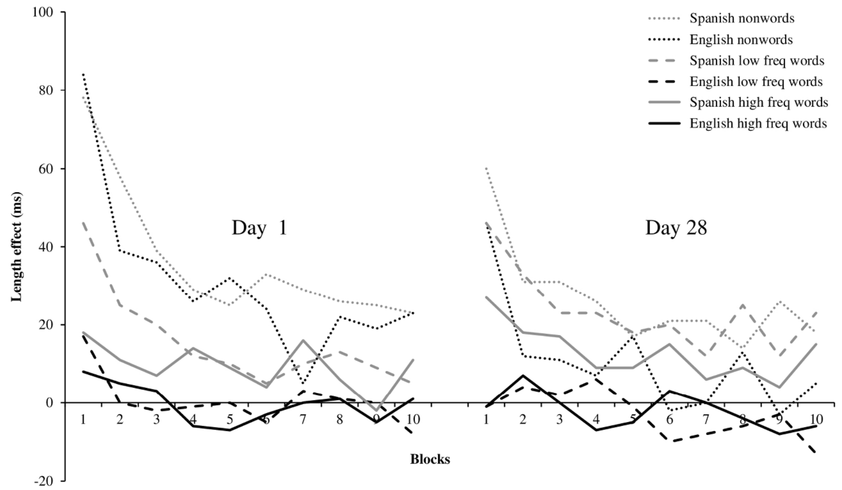

Figure 3. Length effects for high frequency words, low frequency words and nonwords in Experiments 1 (English) and 2 (Spanish).

297x210mm (200 x 200 DPI)